

## A NEW SPECTRAL CONJUGATE GRADIENT ALGORITHM FOR UNCONSTRAINED OPTIMIZATION

KHALIL K. ABBO<sup>1</sup>, YOKSAL A. LAYLANI<sup>2</sup> & HISHAM M KHUDHUR<sup>3</sup>

<sup>1,3</sup>Department of Mathematics, College of computers sciences and mathematics University of Mosul, Mosul, Iraq

<sup>2</sup>Department of Mathematics College of sciences University of Kirkuk, Kirkuk, Iraq

### ABSTRACT

*In this research, we developed a new search direction in the conjugate gradient algorithms by using combined convex property. The developed algorithm becomes converged by assuming some hypothesis. The numerical results show the efficiency of the developed method for solving test unconstrained nonlinear optimization problems.*

**KEYWORDS:** Search Direction, Convergence, .& Optimization Problems

**Received:** Oct 24, 2017; **Accepted:** Nov 14, 2017; **Published:** Dec 29, 2017; **Paper Id.:** IJMCARFEB20181

### 1. INTRODUCTION

Conjugate gradient methods represent an important class of unconstrained optimization algorithms with strong local and global convergence properties and modest memory requirements. An excellent survey of the development of different versions of nonlinear conjugate-gradient methods, with special attention to global convergence properties, is presented by Hager and Zhang [8]. This family of algorithms includes a lot of variants, well known in the literature, with important convergence properties and numerical efficiency. The purpose of this chapter is to present these algorithms as well as their performances to solve a large variety of large-scale unconstrained optimization problems. For solving the nonlinear unconstrained optimization problem

$$\min\{f(x), x \in R^n\} \quad (1)$$

Where  $f: R^n \rightarrow R$  is continuously differentiable function and bounded below, starting from an initial guess  $x \in R^n$  a nonlinear conjugate-gradient method generates a sequence  $\{x_k\}$  as:

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

Where  $\alpha_k > 0$  is obtained by line search, and the directions  $d_k$  are generated as:

$$d_{k+1} = -g_{k+1} + \beta_k s_k, \quad d_1 = -g_1 \quad (3)$$

In (3)  $\beta_k$  is known as the conjugate gradient parameter,  $s_k = x_{k+1} - x_k$  and  $g_k = \nabla f(x_k)$ .

Consider  $\|\cdot\|$  the Euclidean norm and define  $y_k = g_{k+1} - g_k$ . The line search in the conjugate-gradient algorithms is often

based on the standard Wolfe conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq +\rho \alpha_k g_k^T d_k \quad (4)$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \quad (5)$$

Where  $d_k$  is a descent direction and  $0 < \rho \leq \sigma < 1$ . For some conjugate-gradient algorithms, stronger versions of the Wolfe conditions are needed to ensure convergence and to enhance stability. According to the formula for  $\beta_k$  computation, the conjugate-gradient algorithms can be classified as classical, hybrid, scaled, modified and parametric. In the following, we shall present these algorithms and insist on their numerical Dolan and Moré's performances profiles for solving large-scale unconstrained optimization problems. The history of conjugate-gradient method begins with the seminal paper of Hestenes and Stiefel [9], who presented an algorithm for solving symmetric, positive definite linear algebraic systems. In 1964 Fletcher and Reeves [6] extended the domain of application of conjugate-gradient method to nonlinear problems, thus starting the nonlinear conjugate-gradient research direction. The main advantages of the conjugate-gradient method are its low memory requirements and its convergence speed. A large variety of nonlinear conjugate gradient algorithms are known. For each of them, convergence results have been proved in mild conditions which refer to the Lipschitz and boundedness assumptions. To prove the global convergence of nonlinear conjugate-gradient methods, often the Zoutendijk condition is used combined with analysis showing that the sufficient descent condition  $g_k^T d_k \leq -c \|g_k\|^2$  holds, and that there exists a constant  $\delta$  such that  $\|d_k\|^2 \leq \delta k$ . Often, the convergence analysis of conjugate gradient algorithms, for general nonlinear functions, follows insights developed by Gilbert and Nocedal [7]. The idea is to bound the change  $u_{k+1} - u_k$  in the normalized direction  $u_k = d_k / \|d_k\|$ , which is used to conclude, by contradiction, that the gradients cannot be bounded away from zero.

## 2. CLASSICAL CONJUGATE GRADIENT ALGORITHMS

These algorithms are defined by (2) and (3), where the parameter  $\beta_k$  is computed as in Table (1). Observe that these algorithms can be classified as algorithms with  $\|g_{k+1}\|$  in the numerator of  $\beta_k$  and algorithms with  $g_{k+1}^T y_k$  in the numerator of parameter  $\beta_k$ . The FR, CD and DY methods (see the tables for the definitions of the acronyms used for the algorithms throughout the text), with  $\|g_{k+1}\|$  in the numerator of  $\beta_k$  have strong convergence theory, but all performance profiles of conjugate gradient algorithms for unconstrained optimization.

Table 1: Classical Conjugate Gradient Algorithms [1]

No.	(Formula)	(Author)
1.	$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$	(Hestenes-Stiefel (HS), 1952)
2.	$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$	(Fletcher-Reeves (FR), 1964)
3.	$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k}$	(Polak- Ribière (PR), 1969)
4.	$\beta_k^{DX} = -\frac{g_{k+1}^T g_{k+1}}{d_k^T g_k}$	(Dixon (DX), 1975)
5.	$\beta_k^{CD} = -\frac{g_{k+1}^T g_{k+1}}{d_k^T g_k}$	(Fletcher (CD), 1987)
6.	$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k}$	(Liu-Storey (LS), 1991)
7.	$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k}$	(Dai-Yuan (DY), 1999)

These methods are susceptible to jamming. They begin to take small steps without making any significant progress to the minimum. On the other hand, the HS, PRP and LS methods, with  $g_{k+1}^T y_k$  in the numerator of parameter  $\beta_k$ , have a built-in restart feature that addresses the jamming phenomenon. When the step  $s_k$  is small, the factor  $y_k = g_{k+1} - g_k$  in the numerator of  $\beta_k$  tends to zero. Therefore,  $\beta_k$  becomes small and the new direction  $d_{k+1}$  in (3) is essentially the steepest descent direction  $-g_{k+1}$ . With other words, the HS, PRP and LS methods automatically adjust  $\beta_k$  to avoid jamming, and their performances are better than the performance of methods with  $\|g_{k+1}\|$  in the numerator of  $\beta_k$ .

This paper is organized as following. In section 3, new algorithms for CG to solve Unconstrained Optimization Problems. In section, 4 we will show that our algorithm satisfies descent condition for every iteration. Section 5, we will show that our algorithm satisfies Global convergence condition for every iteration. Section 6, presents numerical experiments and comparisons.

### 3. NEW ALGORITHM FOR CG TO SOLVE UNCONSTRAINED OPTIMIZATION PROBLEMS

We can define the new search direction by the following:

$$d_{k+1} = -\theta_k g_{k+1} + (1 - \theta_k) \beta_k d_k \quad (6)$$

$$\theta_k = \frac{|g_{k+1}^T d_k|}{y_k^T d_k} \quad (7)$$

Where  $0 < \theta < 1$

#### New Algorithm

**Step 1:** Initialization. Select  $x_1 \in R^n$  and the parameters  $\theta > 0$ .

Compute  $f(x_1)$  and  $g_1$ . Consider  $d_1 = -g_1$  and set the initial guess  $\alpha_1 = 1 / \|g_1\|$ .

**Step 2:** Test for continuation of iterations. If  $\|g_{k+1}\| \leq 10^{-6}$ , then stop.

**Step 3:** Line search. Compute  $\alpha_{k+1} > 0$  satisfying the Wolfe line search condition (4) and (5) and update the variables  $x_{k+1} = x_k + \alpha_k d_k$ .

**Step 4:** Direction new computation, compute  $d = -\theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k$ ,  $\theta = \frac{|g_{k+1}^T d_k|}{y_k^T d_k}$ . If the restart criterion

$|g_{k+1}^T g_k| \geq 0.2 \|g_{k+1}\|^2$ , is satisfied, then set  $d = -g_{k+1}$  Otherwise define  $d_{k+1} = d$ . Compute the initial step size  $\alpha_k = \alpha_{k-1} \|d_{k-1}\| / \|d_k\|$ , set  $k = k + 1$  and continue with step2.

#### 4. THE DESCENT PROPERTY OF THE NEW METHOD

Below we have to show the descent property for our proposed new Scaled conjugate gradient algorithm, denoted by:

$$d_{k+1} = -\theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k \quad (8)$$

In the following theorem (1).

##### Theorem (1)

The search direction defined by

$$d_{k+1} = -\theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k \quad (9)$$

Satisfies descent property for all  $k \geq 1$

##### Proof

The proof is by induction.

If  $k=1$  then  $d_1^T g_1 = -\|g_1\| < 0$ .

Assume that  $g_k^T d_k < 0$

Note that from the wolfe conditions we have  $d_k^T y_k > 0$ .

We prove that the relation is true when  $k = k + 1$  by multiplying the equation (9) to  $g_{k+1}^T$  we obtain

$$d_{k+1}^T g_{k+1} = -\theta_k g_{k+1}^T g_{k+1} + (1 - \theta_k) \beta_{k+1} g_{k+1}^T d_k \quad (10)$$

$$d_{k+1}^T g_{k+1} = -\frac{|g_{k+1}^T d_k|}{y_k^T d_k} \|g_{k+1}\|^2 + (1 - \frac{|g_{k+1}^T d_k|}{y_k^T d_k}) \beta_{k+1} g_{k+1}^T d_k \quad (11)$$

$$d_{k+1}^T g_{k+1} = -\frac{|g_{k+1}^T d_k|}{y_k^T d_k} \|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k - \frac{|g_{k+1}^T d_k|}{y_k^T d_k} \beta_{k+1} g_{k+1}^T d_k \quad (12)$$

$$\text{Let } \frac{|g_{k+1}^T d_k|}{y_k^T d_k} \|g_{k+1}\|^2 > \beta_{k+1} g_{k+1}^T d_k - \frac{|g_{k+1}^T d_k|}{y_k^T d_k} \beta_{k+1} g_{k+1}^T d_k$$

$$d_{k+1}^T g_{k+1} < 0 \quad (13)$$

## 5. GLOBAL CONVERGENCE ANALYSIS

Next we will show that CG method with  $\beta_{k+1}$  converges globally. We need the following assumption for the convergence of the proposed new algorithm.

### Assumption (1) [4]

1-Assume  $f$  is bound below in the level set  $S = \{x \in R^n : f(x) \leq f(x_0)\}$ ; In some Initial point.

2- $f$  is continuously differentiable and its gradient is Lipschitz continuous, there exist  $L > 0$  such that:

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad \forall x, y \in N \quad (14)$$

3-  $f$  is uniformly convex function, then there exists a constant  $\mu > 0$  such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \|x - y\|^2, \text{ for any } x, y \in S \quad (15)$$

or equivalently

$$y_k^T s_k \geq \mu \|s_k\|^2 \quad \text{and} \quad \mu \|s_k\|^2 \leq y_k^T s_k \leq L \|s_k\|^2 \quad (16)$$

On the other hand, under assumption (1), it is clear that there exist positive constants  $B$  such

$$\|x\| \leq B, \quad \forall x \in S \quad (17)$$

$$\|\nabla f(x)\| \leq \bar{\gamma}, \quad \forall x \in S \quad (18)$$

### Lemma(1) [4,5 and 10]

Suppose that Assumption (1) and equation (17) hold true. Consider any conjugate gradient method in from (2) and

(3), where  $d_k$  is a descent direction and  $\alpha_k$  is obtained by the strong Wolfe line search. If

$$\sum_{k \geq 1} \frac{1}{\|d_{k+1}\|^2} = \infty \quad (19)$$

Then, we have

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0 \quad (20)$$

More details can be found in [3 and 6].

### Theorem (2)

Suppose that assumption (1) and equation (17) and the descent condition hold. Consider a conjugate gradient method in the form

$$d_{k+1} = -\theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k \quad (21)$$

where  $\alpha_k$  is computed from Wolfe line search condition (4) and (5), If the objective function is uniformly on set S, then  $\liminf_{n \rightarrow \infty} \|g_k\| = 0$ .

### Proof

Firstly, we need substituting our  $\beta_{k+1}$  in the direction  $d_{k+1}$  there for we obtain:

$$d_{k+1} = -\theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k \quad (22)$$

After simplify above equation we get

$$\|d_{k+1}\|^2 = \|- \theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k\|^2 \quad (23)$$

$$\|d_{k+1}\|^2 = \|- \theta_k g_{k+1} + (1 - \theta_k) \beta_{k+1} d_k\|^2 \quad (24)$$

$$\|d_{k+1}\|^2 \leq \theta_k \|g_{k+1}\|^2 + \beta_{k+1} \|d_k\|^2 + \theta_k \beta_{k+1} \|d_k\|^2 \quad (25)$$

$$\|d_{k+1}\|^2 \leq \theta_k \|g_{k+1}\|^2 + (1 + \theta_k) \beta_{k+1} \|d_k\|^2 \quad (26)$$

Suppose that  $b = \theta_k$  and  $c = (1 + \theta_k) \beta_{k+1}$

$$\|d_{k+1}\|^2 \leq b \|g_{k+1}\|^2 + c \|d_k\|^2 \quad (27)$$

$$\|d_{k+1}\|^2 \leq b \gamma^2 + c \|d_k\|^2 \quad (28)$$

$$\|d_{k+1}\|^2 \leq \frac{1}{\bar{\gamma}^2} (b(\bar{\gamma}^2)^2 + \bar{\gamma}^2 c \|d_k\|^2) \quad (29)$$

$$\text{Suppose that } T = b(\bar{\gamma}^2)^2 + \bar{\gamma}^2 c \|d_k\|^2$$

$$\|d_{k+1}\|^2 \leq T \frac{1}{\bar{\gamma}^2} \quad (30)$$

$$\sum_{k=1}^{\infty} \frac{1}{\|d_{k+1}\|^2} \leq \frac{1}{T} \bar{\gamma}^{-2} \sum_{k \geq 1} 1 = \infty \quad (31)$$

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad (32)$$

## 6. NUMERICAL RESULTS AND COMPARISONS

In this section, we report some preliminary numerical results. We compare with classical conjugate gradient direction methods. We compare the performance of new formal  $d_{k+1}$  Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization to classical direction conjugate gradient algorithm by using  $\beta_k^{FR}, \beta_k^{HS}, \beta_k^{PRP}, \beta_k^{DY}$ . We have selected (75) large scale unconstrained optimization problem, for each test problems taken from (Andrei, 2008) [2]. For each test function we have considered numerical experiments with the number of variables  $n = 100, \dots, 1000$ . These new versions are compared with well-known classical directionconjugate gradient algorithm. All these algorithms are implemented with standard Wolfe line search conditions (4) and (5) with. In all these cases, the stopping criteria is the  $\|g_k\| = 10^{-6}$ . All codes are written in double precision Fortran Language with F77 default compiler settings. The test functions usually start point standard initially summary numerical results recorded in the figures (1),(2),(3) by matlab. The performance profile by Dolan and More' [5] is used to display the performance of the Proposed new Scaled direction conjugate gradient algorithm with classical directionconjugate gradient algorithm by using  $\beta_k^{FR}, \beta_k^{HS}, \beta_k^{PRP}, \beta_k^{DY}$ . Define  $p = 750$  as the whole set of  $n_p$  test problems and  $S = 8$  the set of the interested solvers. Let  $l_{p,s}$  be the number of objective function evaluations required by solver  $S$  for problem  $p$ . Define the performance ratio as:

$$r_{p,s} = \frac{l_{p,s}}{l_p^*} \quad (33)$$

Where  $l_p^* = \min\{l_{p,s} : s \in S\}$ . It is obvious that  $r_{p,s} \geq 1$  for all  $p, s$ . If a solver fails to solve a problem, the ratio  $r_{p,s}$  is assigned to be a large number  $M$ . The performance profile for each solver  $S$  is defined as the following cumulative distribution function for performance ratio  $r_{p,s}$ ,

$$\rho_s(\tau) = \frac{\text{size}\{p \in P : r_{p,s} \leq \tau\}}{n_p} \quad (34)$$

Obviously,  $p_s(1)$  represents the percentage of problems for which solver  $S$  is the best. See [5] for more details about the performance profile. The performance profile can also be used to analyze the number of iterations, the number of gradient evaluations and the cpu time. Besides, to get a clear observation, we give the horizontal coordinate a log-scale in the following figures.

## Notes

1- By using Wolfe conditions (4) and (5) to choose  $\alpha_k$  [11].

2- Definition of algorithms

$$FRN \Rightarrow d_{k+1} = -\theta_k g_{k+1} + (1-\theta_k)\beta_{k+1}^{FR} d_k$$

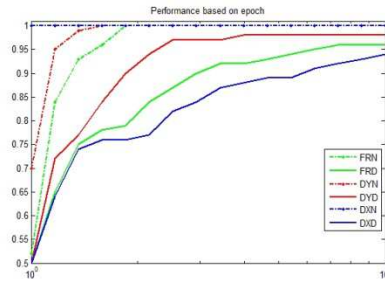
$$FRD \Rightarrow d_{k+1} = -g_{k+1} + \beta_{k+1}^{FR} d_k$$

$$DYN \Rightarrow d_{k+1} = -\theta_k g_{k+1} + (1-\theta_k)\beta_{k+1}^{DY} d_k$$

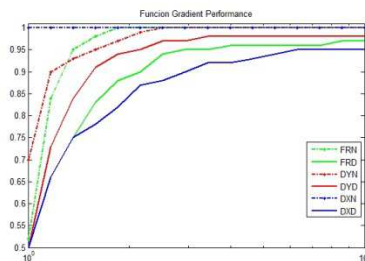
$$DYD \Rightarrow d_{k+1} = -g_{k+1} + \beta_{k+1}^{DY} d_k$$

$$DXN \Rightarrow d_{k+1} = -\theta_k g_{k+1} + (1-\theta_k)\beta_{k+1}^{DX} d_k$$

$$DXD \Rightarrow d_{k+1} = -g_{k+1} + \beta_{k+1}^{DX} d_k$$

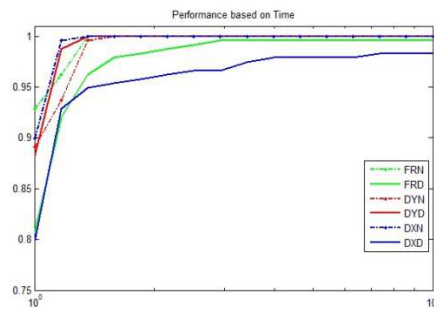


**Figure (1): Performance Based on Iteration**



**Figure (2): Performance Based on Function**





**Figure (3): Performance Based on Time**

## REFERENCES

1. AbboK.(2008), "New CG Method for Large-Scale Unconstrained Optimization Based on Nazareth theorem". *Iraqi Journal of Statistical Sciences* (13) 2008p.p. [53-65].
2. Andrei N. (2008), "An Unconstrained Optimization test function collection". *Adv. Model. Optimization*. 10. pp.147-161.
3. Dai Y. H and Liao L. Z, (2001). *New conjugacy condition and related nonlinear conjugate gradient methods*, *Applied Mathematics and Optimization* 43, pp. 87-101.
4. Rohit Sarin, B. K. Singh & Sanjeev Rajan, *Study on Approximate Gradient Projection (AGP) Property in Nonlinear Programming*, *International Journal of Mathematics and Computer Applications Research (IJMCAR)*, Volume 4, Issue 1, January - February 2014, pp. 19-30
5. Dai Y. H and Yuan Y, (1999). *A nonlinear conjugate gradient method with a strong global convergence property*, *SIAM J. optimization*, pp. 177-182.
6. Dolan. E. D and Mor'e. J. J, "Benchmarking optimization software with performance profiles", *Math. Programming*, 91 (2002), pp. 201-213
7. Fletcher R, Reeves C (1964) *Function minimization by conjugate gradients*. *Comput J* 7:149–154.
8. Gilbert JC, Nocedal J (1992) *Global convergence properties of conjugate gradient methods*. *SIAM J Optim* 2:21–42.
9. Hager WW, Zhang H (2006), *A survey of nonlinear conjugate gradient methods*. *Pac J Optim* 2:35–58.
10. Hestenes MR, Stiefel E (1952) *Methods of conjugate gradients for solving linear systems*. *J Res Nat Bur Stand Sec B*48:409–436.
11. Tomizuka H. and Yabe H, (2004). *A Hybrid Conjugate Gradient method for unconstrained Optimization*
12. Wolfe, M. (1978). "Numerical Methods For Unconstrained Optimization An Introduction". New York: Van Nostrand Reinhold.

